Chapter 76

Sentiment Classification:
Facebook' Statuses Mining in the "Arabic Spring" Era

Jalel Akaichi
University of Tunis, Tunisia

ABSTRACT

In this work, we focus on the application of text mining and sentiment analysis techniques for analyzing Tunisian users' statuses updates on Facebook. We aim to extract useful information, about their sentiment and behavior, especially during the "Arabic spring" era. To achieve this task, we describe a method for sentiment analysis using Support Vector Machine and Naïve Bayes algorithms, and applying a combination of more than two features. The output of this work consists, on one hand, on the construction of a sentiment lexicon based on the Emoticons and Acronyms' lexicons that we developed based on the extracted statuses updates; and on the other hand, it consists on the realization of detailed comparative experiments between the above algorithms by creating a training model for sentiment classification.

INTRODUCTION

Recently, social networks, such as Facebook, Twitter, etc., have taken a significant part of people's lives and activities. Billions of internet users are using social networks not only to stay in touch with their friends, discover new acquaintances and share user-created contents, but also to share their points of views related to variety of subjects through a variety of manners such as wall posting, comments, videos, pictures, etc.

Arab countries count among countries having a huge number of social network users. Social networks become a magic tool for Arab people to promote freedom of speech, human rights and democracy. In 2011, social networks were the spark that makes unexpected revolutions that took place in the Arab countries. At the beginning, it was Tunisia that performed its revolution that lay the country into the "delightful" road of democracy. Tunisia was then the trigger that pushes other countries such as Egypt, Yemen, Syria, etc., to taste the path of "freedom".

On social networks interfaces, people had the opportunities to call for the change, by expressing their sentiments through a multiplicity of posts. Commented images, videos, pictures, articles, etc., were shared on social networks to show dictators' crimes, inequalities between regions, etc.

During the Tunisian revolution, and the rough conditions that they have been enduring, especially during the curfews, Facebook has become the common source of information and one of the most important tools of communication for the Tunisians. From the first of July 2011 until the beginning of December 2012, Facebook gains hundreds of thousands new Tunisian users. At that time, Tunisian Facebook users had tendency to share their feelings, their thoughts and to inform their friends about conditions of their cities or neighborhoods by sharing videos and pictures and especially posting short posts on their walls which was the preferred way to interact with others.

However, there is no previous research that dealt with a novel collection of textual data, which consists of Tunisian Facebook users' statuses, and applied machine learning techniques in order to evaluate their performances on such a dataset.

This work explores the potential applications of text and sentiment mining on statuses updates in order to analyze the Tunisian's behavior during the revolution. For this purpose, we choose a random population having Facebook accounts. It includes males and females, students, workers, housewives, etc. The age of targeted population is varying between 21 and 54 years old.

Through the application of text mining, sentiment analysis techniques and especially machine learning algorithms, we aim to identify the nature of the statuses updates, and to link them to behaviors and sentiments characteristics. This, obviously, will be useful, not only for people that want to know themselves, but also for political decision makers that want to know better their potential electorate.

This chapter is organized as follows. The state of the art and the background of social networks text mining are described in section 2. In section 3, we discuss the methodology of our work and describe the proposed architecture. In section 4, the process of our experiments is described and the results are discussed and evaluated. Finally, section 5 concludes and proposes possible directions for future research.

The expected output of this work consists of the following items:

1. A dataset: There are not any existing data sets of Tunisian Facebook users. Therefore, we create our own dataset which consists of a list of Facebook users (50) and their statuses update (approximately 13 statuses per user).

2. The development of sentiment lexicons: The informal language used on online social networks is a main point to consider before performing any text mining techniques. This is why, we built our special lexicons: Emoticons' lexicon, an Acronyms' lexicon and an Interjections' lexicon.

3. The study of the impact of n-grams ($1 < n < 3$) on analyzing sentiments from Facebook statuses updates. Compared to previous published work, we show that the highest accuracy is reached when we use a combination of unigrams and bigrams as features. We used other features such as part of speech tags and stemming. We do not discuss them in detail due to the lack of significant improvements.

4. The experiments: We study two different machine learning classifiers; Support Vector Machine and Naïve Bayes, which consist on a probabilistic model on the preprocessed data. For experiments, a training model for text polarity was created. Our aim is then to examine which feature set can achieve the highest performance in sentiment classification and which classifiers performs better than the other one.

Related Content

Implementation and Visualization of Conceptual Graphs in CharGer
Harry S. Delugach (2014). *International Journal of Conceptual Structures and Smart Applications (pp. 1-19).*
www.irma-international.org/article/implementation-and-visualization-of-conceptual-graphs-in-charger/134885

UWB Indoor Location for Monitoring Dementia Patients: The Challenges and Perception of a Real-Life Deployment
Agnes Grünerbl, Gernot Bahle, Friedrich Hanserand Paul Lukowicz (2013). *International Journal of Ambient Computing and Intelligence (pp. 45-59).*
www.irma-international.org/article/uwb-indoor-location-for-monitoring-dementia-patients/104160

Advancing Human-Centric Solutions: The Future Trajectory of Soft Computing in Modern Society
Monika Singh T., Kishor Kumar Reddy C.and Kari Lippert (2024). *Human-Machine Collaboration and Emotional Intelligence in Industry 5.0 (pp. 1-18).*
www.irma-international.org/chapter/advancing-human-centric-solutions/351553

Fuzzy Logic for Breast Cancer Diagnosis Using Medical Thermogram Images
Surekha Kamath (2015). *Fuzzy Expert Systems for Disease Diagnosis (pp. 168-199).*
www.irma-international.org/chapter/fuzzy-logic-for-breast-cancer-diagnosis-using-medical-thermogram-images/124447

Emerging Trends in Edge AI for Industry 4.0 and 5.0: Technologies, Applications, and Challenges
S. Anusha, M. Ezhilvandan, S. Gayathri, V. Sivaraman, Anish T. P.and Siva Subramanian R. (2024). *AI Innovations in Service and Tourism Marketing (pp. 211-228).*
www.irma-international.org/chapter/emerging-trends-in-edge-ai-for-industry-40-and-50/352830